

Applications of Deep Learning in Molecule Generation and Molecular Property Prediction.

Year: 2020 | Citations: 290 | Authors: W. Patrick Walters, R. Barzilay

Abstract

Recent advances in computer hardware and software have led to a revolution in deep neural networks that has impacted fields ranging from language translation to computer vision. Deep learning has also impacted a number of areas in drug discovery, including the analysis of cellular images and the design of novel routes for the synthesis of organic molecules. While work in these areas has been impactful, a complete review of the applications of deep learning in drug discovery would be beyond the scope of a single Account. In this Account, we will focus on two key areas where deep learning has impacted molecular design: the prediction of molecular properties and the de novo generation of suggestions for new molecules. One of the most significant advances in the development of quantitative structure-activity relationships (QSARs) has come from the application of deep learning methods to the prediction of the biological activity and physical properties of molecules in drug discovery programs. Rather than employing the expert-derived chemical features typically used to build predictive models, researchers are now using deep learning to develop novel molecular representations. These representations, coupled with the ability of deep neural networks to uncover complex, nonlinear relationships, have led to state-of-the-art performance. While deep learning has changed the way that many researchers approach QSARs, it is not a panacea. As with any other machine learning task, the design of predictive models is dependent on the quality, quantity, and relevance of available data. Seemingly fundamental issues, such as optimal methods for creating a training set, are still open questions for the field. Another critical area that is still the subject of multiple research efforts is the development of methods for assessing the confidence in a model. Deep learning has also contributed to a renaissance in the application of de novo molecule generation. Rather than relying on manually defined heuristics, deep learning methods learn to generate new molecules based on sets of existing molecules. Techniques that were originally developed for areas such as image generation and language translation have been adapted to the generation of molecules. These deep learning methods have been coupled with the predictive models described above and are being used to generate new molecules with specific predicted biological activity profiles. While these generative algorithms appear promising, there have been only a few reports on the synthesis and testing of molecules based on designs proposed by generative models. The evaluation of the diversity, quality, and ultimate value of molecules produced by generative models is still an open question. While the field has produced a number of benchmarks, it has yet to agree on how one should ultimately assess molecules "invented" by an algorithm.